

SARS-CoV-2 Delta–Omicron Recombinant Viruses, United States

Kristine A. Lacek, Benjamin L. Rambo-Martin, Dhvani Batra, Xiao-yu Zheng, Norman Hassell, Hitoshi Sakaguchi, Thomas Peacock, Natalie Groves, Matthew Keller, Malania M. Wilson, Mili Sheth, Morgan L. Davis, Mark Borroughs, Jonathan Gerhart, Samuel S. Shepard, Peter W. Cook, Justin Lee, David E. Wentworth, John R. Barnes, Rebecca Kondor, Clinton R. Paden

To detect new and changing SARS-CoV-2 variants, we investigated candidate Delta–Omicron recombinant genomes from Centers for Disease Control and Prevention national genomic surveillance. Laboratory and bioinformatic investigations identified and validated 9 genetically related SARS-CoV-2 viruses with a hybrid Delta–Omicron spike protein.

Emerging variants of SARS-CoV-2 are characterized and monitored closely by national genomic surveillance. In addition to sequencing efforts from US public health, academic, and commercial laboratories, the Centers for Disease Control and Prevention (CDC) collects and sequences SARS-CoV-2 specimens from 64 partners across state, tribal, local, and territorial public health agencies through the National SARS-CoV-2 Strain Surveillance program (<https://www.cdc.gov/coronavirus/2019-ncov/variants/cdc-role-surveillance.html>) and funds SARS-CoV-2 sequencing through a nationwide network of commercial laboratory testing companies. To date, these efforts have contributed 1.8 million SARS-CoV-2 genomes from the United States to public repositories. The purpose of this genomic surveillance system is to detect and respond dynamically to new and changing SARS-CoV-2 variants (1).

Recombination is an evolutionary mechanism frequently observed in coronaviruses (2,3), and it can lead to rapid accumulation of mutations and heightened transmissibility (4). SARS-CoV-2 recombination events have also been found to arise disproportionately in the spike gene (Y. Turkahia et al., unpub. data, <https://www.biorxiv.org/content/10.1101/2021.08.04.455157V1>). Recombination between Alpha and Delta SARS-CoV-2 variants has been documented (5–7).

Given the divergence of the Delta and Omicron variant genomes, as well as the known immune-escape properties of Omicron (8,9), a Delta–Omicron recombinant strain could alter the landscape of vaccine and therapeutic effectiveness. In early 2022, viruses resulting from recombination between Delta and Omicron were reported, but further inspection indicated that these claims seemed to have resulted from laboratory artifact or co-infections (10). With this study, we identified candidate Delta–Omicron recombinant genomes from the CDC national genomic surveillance and attempted to rule out laboratory contamination or sequencing error.

The Study

We identified 9 candidate recombinant sequences (Table) from CDC national genomic surveillance dataset made publicly available in GenBank and GISAID EpiCoV (<https://www.gisaid.org>). Using Bolotie, a rapid interclade recombination detection method (3), we identified these sequences as candidate recombinant genomes, having 1 parent in Delta (clade 21J) and 1 in Omicron (clade 21K). Bolotie describes a single breakpoint between nucleotide positions 22035 and 22577 (referenced to GenBank accession no. NC_045512.2); there are no differentiating mutations between clades 21J and 21K within this range. These sequences (EPI_ISL_8720194, EPI_

Author affiliations: Centers for Disease Control and Prevention, Atlanta, Georgia, USA (K.A. Lacek, B.L. Rambo-Martin, D. Batra, X. Zheng, N. Hassell, M. Keller, M.M. Wilson, M. Sheth, M.L. Davis, M. Burroughs, J. Gerhart, S.S. Shepard, P.W. Cook, J. Lee, D.E. Wentworth, J.R. Barnes, R. Kondor, C.R. Paden); General Dynamics Information Technology, Inc., Atlanta (X. Zheng); Retired, Yotsukaido City, Japan (H. Sakaguchi); UK Health Security Agency, London, UK (T. Peacock, N. Groves); Imperial College London, London (T. Peacock); ASRT Incorporated, Smyrna, Georgia, USA (M.L. Davis, J. Gerhart)

DOI: <https://doi.org/10.3201/eid2807.220526>

Table. Candidate recombinant samples, states, collection dates, and Bolotie outputs for the SARS-CoV-2 AY.119.2:BA.1.1 recombinant cluster, United States*

| GISAID accession no. | State | GISAID virus name | Collection date | Bolotie results |
|----------------------|-------|-----------------------------------------|-----------------|--------------------------------------------------|
| EPI_ISL_8720194 | TN | hCoV-19/USA/TN-CDC-ASC210559252/2021 | 2021 Dec 31 | 21J (Delta): 1-22032; 21K (Omicron): 22033-29903 |
| EPI_ISL_9147438 | NJ | hCoV-19/USA/NJ-CDC-IBX952397337138/2022 | 2022 Jan 4 | 21J (Delta): 1-22032; 21K (Omicron): 22033-29903 |
| EPI_ISL_8981712 | PA | hCoV-19/USA/PA-CDC-LC0473996/2022 | 2022 Jan 4 | 21J (Delta): 1-22032; 21K (Omicron): 22033-29903 |
| EPI_ISL_8981824 | PA | hCoV-19/USA/PA-CDC-LC0474055/2022 | 2022 Jan 4 | 21J (Delta): 1-22032; 21K (Omicron): 22033-29903 |
| EPI_ISL_8981459 | PA | hCoV-19/USA/PA-CDC-LC0474301/2022 | 2022 Jan 4 | 21J (Delta): 1-22032; 21K (Omicron): 22033-29903 |
| EPI_ISL_9088187 | MA | hCoV-19/USA/MA-CDC-STM-HZEBR92XC/2022 | 2022 Jan 12 | 21J (Delta): 1-22032; 21K (Omicron): 22033-29903 |
| EPI_ISL_9147935 | NJ | hCoV-19/USA/NJ-CDC-IBX640654818289/2022 | 2022 Jan 12 | 21J (Delta): 1-22032; 21K (Omicron): 22033-29903 |
| EPI_ISL_10389336 | NJ | hCoV-19/USA/NJ-CDC-ASC210553977/2022 | 2022 Feb 12 | 21J (Delta): 1-22032; 21K (Omicron): 22033-29903 |
| EPI_ISL_10389339 | NJ | hCoV-19/USA/NJ-CDC-ASC210553978/2022 | 2022 Feb 12 | 21J (Delta): 1-22032; 21K (Omicron): 22033-29903 |

*These 9 candidate recombinant viruses were identified by an exhaustive search of publicly available SARS-CoV-2 viral genomes with orf1ab:2855V, 4176N, 6248S and S:95I, 142D, 157-, 346K, 501Y mutations. hCoV-19/USA/PA-CDC-LC0474055/2022 and hCoV-19/USA/PA-CDC-LC0474301/2022 underwent resequencing at the Centers for Disease Control and Prevention. Bolotie (3) identified all 9 as recombinant genomes between Delta (clade 21J) and Omicron (clade 21K). Bolotie cannot determine the true breakpoint because of high sequence homology, but the same region is identified for all 9 sequences (nt position 22032 as referenced to GenBank accession no. NC_045512.2).

ISL_9147438, EPI_ISL_9147935, EPI_ISL_8981459, EPI_ISL_8981824, EPI_ISL_9088187 [A. Bolze et al., unpub. data, <https://www.medrxiv.org/content/medrxiv/early/2022/03/12/2022.03.09.22272113.full.pdf>], EPI_ISL_8981712, EPI_ISL_10389339, EPI_ISL_10389336) contain hallmark mutation sets from both Omicron and Delta SARS-CoV-2 lineages, changing from Delta-associated substitutions to Omicron-associated substitutions between spike protein amino acids 158 and 339 (Appendix Figure 1, panel A, <https://wwwnc.cdc.gov/EID/article/28/7/22-0526-App1.pdf>). This breakpoint is distinct from the 2 clusters of apparent Delta–Omicron recombinants identified in the United Kingdom (<https://github.com/cov-lineages/pango-designation/issues/422> and <https://github.com/cov-lineages/pango-designation/issues/441>), which have a breakpoint upstream of spike in the *ORF1ab* gene (Appendix Figure 1, panel A), and these samples show a singular breakpoint, unlike concurrently observed Delta–Omicron recombinants in France (P. Colson et al., unpub. data, <https://www.medrxiv.org/content/10.1101/2022.03.03.22271812V1>).

To rule out Delta and Omicron co-infection, laboratory contamination, and bioinformatic error, we examined the raw read data from the 9 candidate recombinants created from molecular loop and amplicon-based sequencing strategies. Two of these specimens were readily available from the original diagnostic laboratory, and extracted RNA was shipped to CDC for confirmatory sequencing. We used Illumina (<https://www.illumina.com>) and PacBio

(<https://www.pacb.com>) sequencing of 2 whole-genome amplicon strategies, as well as spike-gene amplification followed by Oxford Nanopore (<https://nanoporetech.com>) sequencing (Appendix). All sequencing strategies yielded functionally identical consensus sequences compared with the corresponding original sequencing strategies.

Nextclade (11) classified the 9 whole genomes as 21K (Omicron/BA.1). We then split the genomes at position 22150 (within the predicted recombination site range). Nextclade classified the first 22150 base fragment as clade 21J (Delta) and the remainder as clade 21K (Omicron/BA.1). Pangolin version 3.1.20 (pangoLEARN 1.2.123, Scorpio 0.3.16, <https://cov-lineages.org>) assigned a lineage of none to the full-genome sequences. Pangolin classified the first 22150 base fragment of each recombinant as AY.43 (Delta), although the call was not supported by Scorpio. Inspection of this region revealed closer homology to AY.119.2 (Delta) sequences because of mutations orf1ab:A2855V and orf1ab:A6248S, which are common to AY.119 lineages, and orf1ab:K4176N, which is found in a subset of AY.119.2 (Delta) sequences. The remaining sequence fragment from nt 22151 to the 3' end was classified by pangolin as BA.1.1 (Omicron). This observation has been documented in the PANGO-designations repository (<https://github.com/cov-lineages/pango-designation/issues/439>) and is under review for potential lineage assignment.

Detailed sequence analysis confirmed the 2 resequenced specimens as true recombinants and indicated no evidence of co-infection or contamination.

Comparison with a representative AY.119.2 (Delta) specimen indicated characteristic Delta mutations (C21618G, C21846T, G21987A, and deletion 22029–22034) at >99% frequency (>600× coverage for Oxford Nanopore, >1,800× coverage for PacBio, >1000× coverage for Illumina) in the 5' end of the recombinant (Appendix Figure 1, panel B). The 2 BA.1.1 (Omicron) deletions at the beginning of the spike gene (21765–21770 and 21987–21995) and the characteristic Omicron 9-base insertion after nt 22205 were not present in read data, consistent with a Delta origin for the 5' end of the spike gene. After position 22577, the mutation profiles mirrored that of a representative BA.1.1 (Omicron) specimen (Appendix Figure 1, panel B). Analysis of individual Oxford Nanopore reads showed characteristic Delta mutations co-occurring with Omicron single-nucleotide variants on the same reads (sharing Delta 22029–22034 deletion and Omicron 22673 T>C; Appendix Figure 2). The translated spike protein is a hybrid, containing characteristic amino acids from both Delta and Omicron parents with a breakpoint between the N terminal domain and receptor-binding domain of spike S1 protein (Appendix Figure 1, panel A).

To visualize the parents of the recombinant genomes, we split all candidate recombinant genomes at position 22150, within the predicted breakpoint, and used Nextclade (11) to place each genome fragment (1–22150 and 22151 through the 3' end) onto a reference tree. We visualized the 2 trees as a tanglegram tree with Auspice (12). Nucleotides 1–22150 clustered with clade 21J (Delta) sequences, and the remaining fragment of the genome clustered with 21K (Omicron/BA.1) (Appendix Figure 3).

Conclusions

Our results provide evidence of a recombinant SARS-CoV-2 genome containing a hybrid spike protein derived from a Delta (AY.119.2)–Omicron (BA.1.1) recombination event. However, the ability to effectively identify and confirm additional recombinant viruses remains challenging because of the range of sequence quality available in the public domain. These limitations are a result of amplification inefficiency and consensus-calling algorithmic error, as well as cases of co-infection or potential sample contamination.

Comparative phenotypic characterization of virus isolates from the recombinant cluster was not possible because all specimens were chemically inactivated. In the spike protein, there are no additional amino acid substitutions within the receptor-binding domain compared with BA.1.1 (Omicron) lineage viruses. Recombinant viruses with this hybrid spike protein were

detected over the course of 6 weeks, but the number of cases resulting from those viruses remains low. Most cases were identified within the mid-Atlantic region of the United States. However, epidemiologic linkage cannot be determined because CDC does not collect identifying information for these samples.

Systematic virus surveillance is essential for long-term monitoring of SARS-CoV-2 evolution. Given the potential public health consequences of new variants emerging from recombination, investigations involving laboratory and bioinformatic components, such as the one presented here, are critical for correctly identifying and tracking these viruses.

Acknowledgments

We thank the public health program and laboratory staff members who contribute to National SARS-CoV-2 Strain Surveillance, including the Association of Public Health Laboratories, commercial laboratory staff members, laboratories submitting tagged baseline SARS-CoV-2 sequences via GISAID EpiCov and GenBank, CDC COVID-19 Pandemic Response Laboratory Task Force, Strain Surveillance and Emerging Variants Bioinformatics Working Group, Natalie Thornburg, and colleagues in the UK Health Security Agency. Furthermore, we gratefully acknowledge the authors from the originating laboratories responsible for obtaining the specimens and the submitting laboratories where genetic sequence data were generated and shared via GISAID. A full acknowledgment table can be found with the EPI_SET ID EPI_SET_20220412tv, in the Data Acknowledgement Locator under GISAID resources (<https://www.gisaid.org/epi-search>).

All authors have completed and submitted the International Committee of Medical Journal Editors form for disclosure of potential conflicts of interest. No potential conflicts of interest were disclosed.

About the Author

Ms. Lacey is an informatics health scientist with the Influenza Division, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia. Her primary research interests include viral genomics, next-generation sequencing, and genome assembly.

References

1. National Center for Immunization and Respiratory Diseases (NCIRD) DoVD. Science brief: emerging SARS-CoV-2 variants [cited 2022 Feb 17]. <https://www.cdc.gov/coronavirus/2019-ncov/science/science-briefs/scientific-brief-emerging-variants.html>

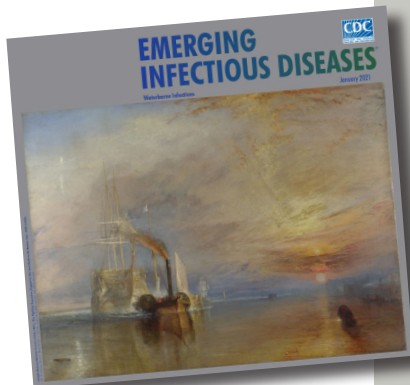
2. Lai MM, Baric RS, Makino S, Keck JG, Egbert J, Leibowitz JL, et al. Recombination between nonsegmented RNA genomes of murine coronaviruses. *J Virol*. 1985;56:449–56. <https://doi.org/10.1128/jvi.56.2.449-456.1985>
3. Varabyou A, Pockrandt C, Salzberg SL, Pertea M. Rapid detection of inter-clade recombination in SARS-CoV-2 with Bolotie. *Genetics*. 2021;218:iyab074.
4. Li X, Giorgi EE, Marichannegowda MH, Foley B, Xiao C, Kong X-P, et al. Emergence of SARS-CoV-2 through recombination and strong purifying selection. *Sci Adv*. 2020;6:eabb9153. <https://doi.org/10.1126/sciadv.abb9153>
5. Sekizuka T, Itokawa K, Saito M, Shimatani M, Matsuyama S, Hasegawa H, et al. Genome recombination between Delta and Alpha variants of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). *Jpn J Infect Dis*. 2022 Feb 28 [Epub ahead of print].
6. Escalera-Zamudio M, Sánchez HGC, Arredondo LJD, Gutiérrez BRE. Proposal to redesignate B.1.631 as recombinant lineage XB [cited 2022 Feb 18]. <https://virological.org/t/re-proposal-to-redesignate-b-1-631-as-recombinant-lineage-xb/746>
7. Jackson B, Rambaut A, Pybus OG, Robertson DL, Connor T, Loman NJ, et al. Recombinant SARS-CoV-2 genomes involving lineage B.1.1.7 in the UK [cited 2022 Feb 18]. <https://virological.org/t/recombinant-sars-cov-2-genomes-involving-lineage-b-1-1-7-in-the-uk/658>
8. Liu L, Iketani S, Guo Y, Chan JF, Wang M, Liu L, et al. Striking antibody evasion manifested by the Omicron variant of SARS-CoV-2. *Nature*. 2022;602:676–81. <https://doi.org/10.1038/s41586-021-04388-0>
9. Chakraborty C, Bhattacharya M, Sharma AR. Emerging mutations in the SARS-CoV-2 variants and their role in antibody escape to small molecule-based therapeutic resistance. *Curr Opin Pharmacol*. 2022;62:64–73. <https://doi.org/10.1016/j.coph.2021.11.006>
10. Kreier F. Deltacron: the story of the variant that wasn't. *Nature*. 2022;602:19. <https://doi.org/10.1038/d41586-022-00149-9>
11. Aksamentov I, Roemer C, Hodcroft EB, Neher RA. Nextclade: clade assignment, mutation calling and quality control for viral genomes. *J Open Source Softw*. 2021;6:3773. <https://doi.org/10.21105/joss.03773>
12. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, et al. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics*. 2018;34:4121–3. <https://doi.org/10.1093/bioinformatics/bty407>

Address for correspondence: Clinton R. Paden, Centers for Disease Control and Prevention, 1600 Clifton Rd NE, Mailstop H18-6, Atlanta, GA 30329-4027, USA; email: cpaden@cdc.gov

etymologia revisited

Petri Dish

[pe'tre 'dish]



**Originally published
in January 2021**

The Petri dish is named after the German inventor and bacteriologist Julius Richard Petri (1852–1921). In 1887, as an assistant to fellow German physician and pioneering microbiologist Robert Koch (1843–1910), Petri published a paper titled “A minor modification of the plating technique of Koch.” This seemingly modest improvement (a slightly larger glass lid), Petri explained, reduced contamination from airborne germs in comparison with Koch’s bell jar.

Sources:

1. Central Sheet for Bacteriology and Parasite Science [in German]. Biodiversity Heritage Library. Volume 1, 1887 [cited 2020 Aug 25]. <https://www.biodiversitylibrary.org/item/210666#page/313/mode/1up>
2. Petri JR. A minor modification of the plating technique of Koch [in German]. *Cent für Bacteriol und Parasitenkd*. 1887;1:279–80.
3. Shama G. The “Petri” dish: a case of simultaneous invention in bacteriology. *Endeavour*. 2019;43:11–6. DOI: [10.1016/j.endeavour.2019.04.001](https://doi.org/10.1016/j.endeavour.2019.04.001)
4. The big story: the Petri dish. *The Biomedical Scientist*. Institute of Biomedical Science [cited 2020 Aug 25]. <https://thebiomedicalscientist.net/science/big-story-petri-dish>

https://wwwnc.cdc.gov/eid/article/27/1/et-2701_article

SARS-CoV-2 Delta–Omicron Recombinant Viruses, United States

Appendix

Supplemental Methods

Sample selection

We aligned candidate recombinant consensus sequences against representative sequences for Delta (AY.4, GISAID accession EPI_ISL_1758376) and Omicron (BA.1, GISAID accession EPI_ISL_7569944) and calculated Hamming distances, sequence quality, and locations of amino acid substitutions.

We also performed an exhaustive search of publicly available SARS-CoV-2 viral genomes with orf1ab:2855V,4176N,6248S and S:95I,142D,157-,346K,501Y variants to identify other potential AY.119:BA.1 recombinant genome sequences from publicly available sequence data in GISAID and NCBI's GenBank.

All samples identified were collected and sequenced as part of CDC's national genomic surveillance activity (1,2), also described here: <https://www.cdc.gov/coronavirus/2019-ncov/variants/cdc-role-surveillance.html>.

We found two complete samples matching the mutation profile for Delta (AY.119.2) until amino acid 339, then matched BA.1.1 (Omicron) through the end of the genome. Examination of raw read alignments showed no evidence of minor variant populations indicative of co-infection. We selected these two samples for additional sequencing via Pacific Biosciences, Oxford Nanopore Technologies, and Illumina platforms due to immediate availability, and high quality of the initial sequence data.

Sequencing Methods

We completed targeted Nanopore Spike gene sequencing with a custom primer set producing 2.2 and 2.5 kb amplicons overlapping across the S1-S2 cleavage site and extending beyond the coding region of the S-gene. Amplicons underwent native barcoding (LSK-109), and

GridION sequencing (Keller et al, in preparation). Live super-accurate basecalling was performed with guppy version 5.1.13 and reads were trimmed with BBDuk version 38.84. We assembled consensus genomes using the Iterative Refinement Meta-Assembler (IRMA) (3) on the CoV-s-gene configuration. Deletion thresholds required 10x depth with >50% frequency.

We prepared the samples using HiFiViral SARS-CoV-2 kit for sequencing on PacBio Sequel II and generated HiFi reads using PacBio SMRT Link v10.2.0.133424. Trimmed HiFi reads were assembled using IRMA with the CoV-pacbio configuration.

Finally, we performed 2x150 Illumina sequencing following library preparation with the IDT xGen SARS-CoV-2 Amplicon Panel (formerly Swift Normalase Amplicon Panel for SARS-CoV-2), down-sampled to 1 million reads per sample, trimmed primers with BBDuk, and assembled consensus genomes with IRMA's default CoV configuration.

We performed clade assignments using Nextclade version 1.13.2 (4) and assigned lineages using Pangolin version 3.1.20 (pangoLEARN 1.2.123, Scorpio 0.3.16) (5) for the whole genome, as well as upstream and downstream of the suspected recombination site independently.

Delta-Omicron Allele Profiles

We calculated the frequency of amino acid residues at each position along the genome within each PANGO lineage for all publicly available data. An allele was considered strongly “Delta” (leftmost orange in Appendix Figure 1) if the residue appeared in more than 50% of sequences in the Delta lineages (B.1.617.2 and AY sublineages), less than 50% of sequences in Omicron lineages (B.1.1.529, BA.1, BA.1.1, BA.2, and BA.3), and its frequency in Delta lineages was more than 20% greater than other lineages (non-Omicron, non-Delta). A residue was considered strongly Omicron (rightmost violet in Appendix Figure 1) if it appeared in more than 50% Omicron sequences, less than 50% of Delta sequences and its frequency in Omicron lineages was more than 20% greater than in other lineages. A residue that was otherwise Omicron, but greater than 20% of Delta sequences contain the residue, it is noted as “Omicron, minor Delta.” Mutant residues that comprise >50% of both Omicron and Delta (common mutations to both) are colored grey. Residues more common in Omicron than Delta and comprising >0.25% Omicron are lighter hues of violet,” and residues more common in Delta lineages than Omicron lineages and comprising >0.25% Delta are colored in lighter hues of orange. White colored alleles match the reference genome (Wuhan-Hu-1), and black colored

alleles are common to neither Omicron nor Delta genomes (although these may be common to a particular sublineage).

Recombination Analysis with Bolotie

Bolotie source code was obtained and installed (6). To build the conditional probability table for Bolotie, 226,173 SARS-CoV-2 genome sequences were selected and aligned to the NC_045512.2 genome. These sequences were published by CDC to GISAID before March 10, 2022, and each sequence contained less than 1% non-ATCG calls. Clade assignments for these sequences were obtained using Nextclade. Based on the phylogenetic relationships of Nextstrain clades, the clades were further binned into 10 large groups (1: 19B, 20A, 20B, 21B, 21D, 21H, 20C; 2: 20D, 21G, 20J, 21E; 3: 20G, 20H, 21C; 4: 21F; 5: 20I; 6: 21A; 7: 21I; 8: 21J (Delta); 9: 21K(Omicron); 10: 21L). A conditional probability table was built with this dataset and grouping using Bolotie default settings. The nine sequences of interest were then analyzed using this conditional probability table.

Phylogenomic Tree Generation

To obtain a reference sample set for the phylogenetic validation and comparison of recombinant samples, all available data (Internal CDC, GISAID, NCBI) was subset in the following fashion:

- Using Nextclade (v1.11.0) (4) clade assignments, five strain sequences from each non-Omicron and non-Delta clade were randomly sampled.
- 53 randomly sampled Delta variant strain sequences.
- 61 randomly sampled Omicron variant strain sequences.
- The Wuhan/Hu-1/2019 reference (to serve as a root sequence).

All reference samples were required to pass the following QC metrics: Nextclade “qc_overallstatus” = “good”, Nextclade “qc_missingdata_status” = “good”, Nextclade “privatenucmutations_totalprivatesubstitutions” < 10, Nextclade “privatenucmutations_totalreversionsubstitutions” < 1, 98% coverage of the spike protein coding frame, and no more than two ambiguous codes within the spike protein coding frame.

From this reference sequence set, two separate datasets were generated, by adding the “start” (positions 1-22150) and “end” (22150 onwards) portion of putative recombinant sample

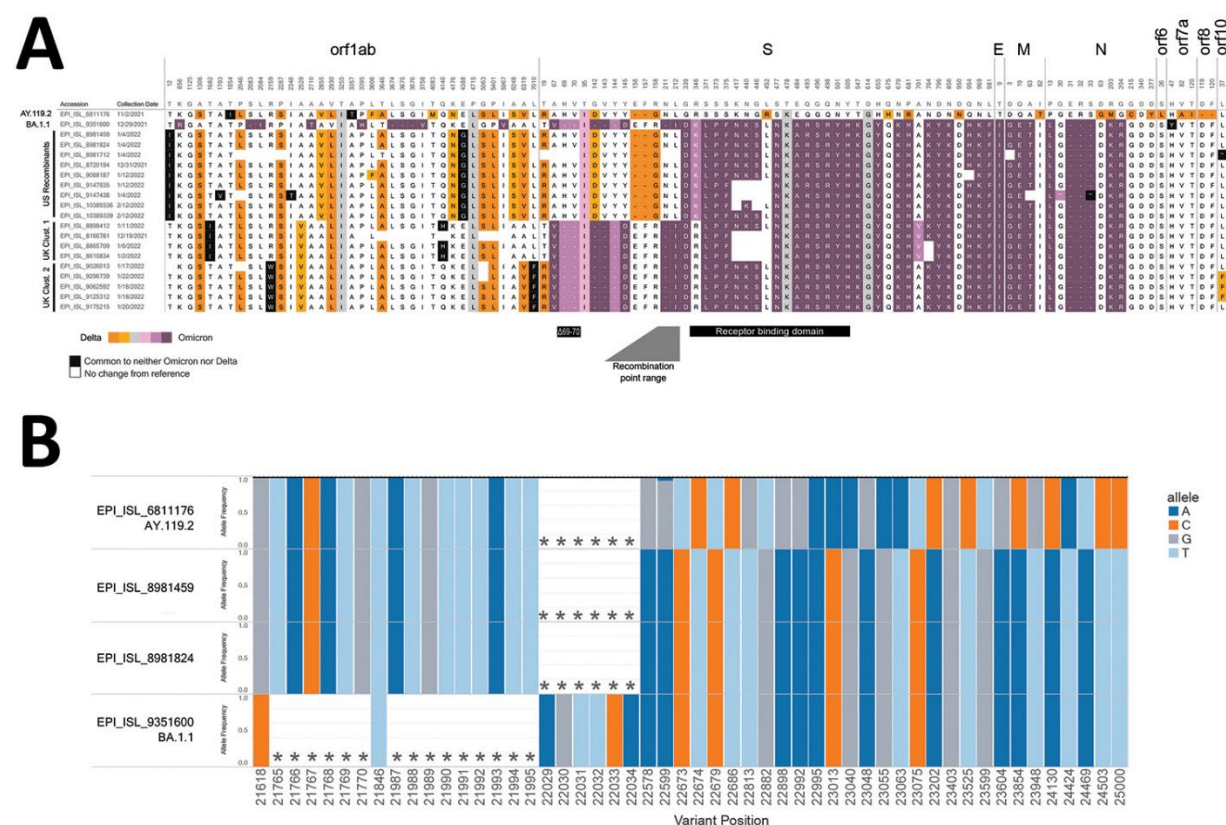
genomes to the references. Two phylogenetic trees were generated using the Nextstrain bioinformatics toolkit (7) with custom “ncov” build profiles (<https://github.com/nextstrain/ncov>) for the “start” and “end” datasets. These trees were then visualized together in Nextstrain’s auspice tree viewer with a tree “tangle-gram” to clearly indicate phylogenetic placement of the recombinant samples on the “start” and “end” dataset trees.

References

1. Lambrou AS, Shirk P, Steele MK, Paul P, Paden CR, Cadwell B, et al.; Strain Surveillance and Emerging Variants Bioinformatic Working Group; Strain Surveillance and Emerging Variants NS3 Working Group. Genomic surveillance for SARS-CoV-2 variants: predominance of the Delta (B.1.617.2) and Omicron (B.1.1.529) variants—United States, June 2021–January 2022. *MMWR Morb Mortal Wkly Rep.* 2022;71:206–11. [PubMed](#) <https://doi.org/10.15585/mmwr.mm7106a4>
2. Paul P, France AM, Aoki Y, Batra D, Biggerstaff M, Dugan V, et al. Genomic surveillance for SARS-CoV-2 variants circulating in the United States, December 2020–May 2021. *MMWR Morb Mortal Wkly Rep.* 2021;70:846–50. [PubMed](#) <https://doi.org/10.15585/mmwr.mm7023a3>
3. Shepard SS, Meno S, Bahl J, Wilson MM, Barnes J, Neuhaus E. Viral deep sequencing needs an adaptive approach: IRMA, the iterative refinement meta-assembler. *BMC Genomics.* 2016;17:708. [PubMed](#) <https://doi.org/10.1186/s12864-016-3030-6>
4. Aksamentov I, Roemer C, Hodcroft EB, Neher RA. Nextclade: clade assignment, mutation calling and quality control for viral genomes. *J Open Source Softw.* 2021;6:3773. <https://doi.org/10.21105/joss.03773>
5. O’Toole Á, Scher E, Underwood A, Jackson B, Hill V, McCrone JT, et al. Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. *Virus Evol.* 2021;7:b064. [PubMed](#) <https://doi.org/10.1093/ve/veab064>
6. Varabyou A, Pockrandt C, Salzberg SL, Pertea M. Rapid detection of inter-clade recombination in SARS-CoV-2 with Bolotie. *Genetics.* 2021;218:iyab074. [PubMed](#) <https://doi.org/10.1093/genetics/iyab074>
7. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, et al. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics.* 2018;34:4121–3. [PubMed](#) <https://doi.org/10.1093/bioinformatics/bty407>

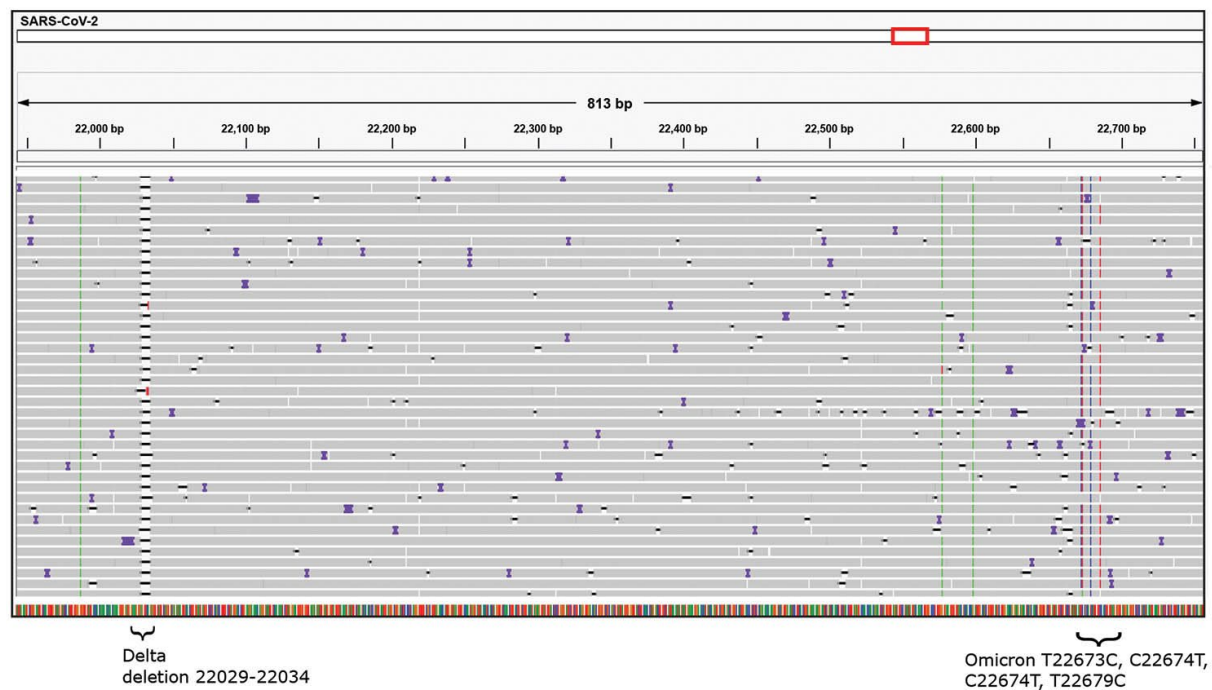
Appendix Table. Data availability

| GISAI Accession | GISAI Virus Name | GenBank Accession | BioSample Accession |
|------------------|-----------------------------------------|-------------------|---------------------|
| EPI_ISL_9088187 | hCoV-19/USA/MA-CDC-STM-HZE92XC/2022 | OM372355 | SAMN25233756 |
| EPI_ISL_10389336 | hCoV-19/USA/NJ-CDC-ASC210553977/2022 | OM835119 | SAMN26256159 |
| EPI_ISL_10389339 | hCoV-19/USA/NJ-CDC-ASC210553978/2022 | OM835120 | SAMN26256158 |
| EPI_ISL_9147935 | hCoV-19/USA/NJ-CDC-IBX640654818289/2022 | OM393452 | SAMN25262455 |
| EPI_ISL_9147438 | hCoV-19/USA/NJ-CDC-IBX952397337138/2022 | OM392955 | SAMN25262195 |
| EPI_ISL_8981712 | hCoV-19/USA/PA-CDC-LC0473996/2022 | OM344166 | SAMN25170131 |
| EPI_ISL_8981824 | hCoV-19/USA/PA-CDC-LC0474055/2022 | OM344238 | SAMN25170117 |
| EPI_ISL_8981459 | hCoV-19/USA/PA-CDC-LC0474301/2022 | OM344043 | SAMN25170048 |
| EPI_ISL_8720194 | hCoV-19/USA/TN-CDC-ASC210559252/2021 | OM272834 | SAMN24974483 |

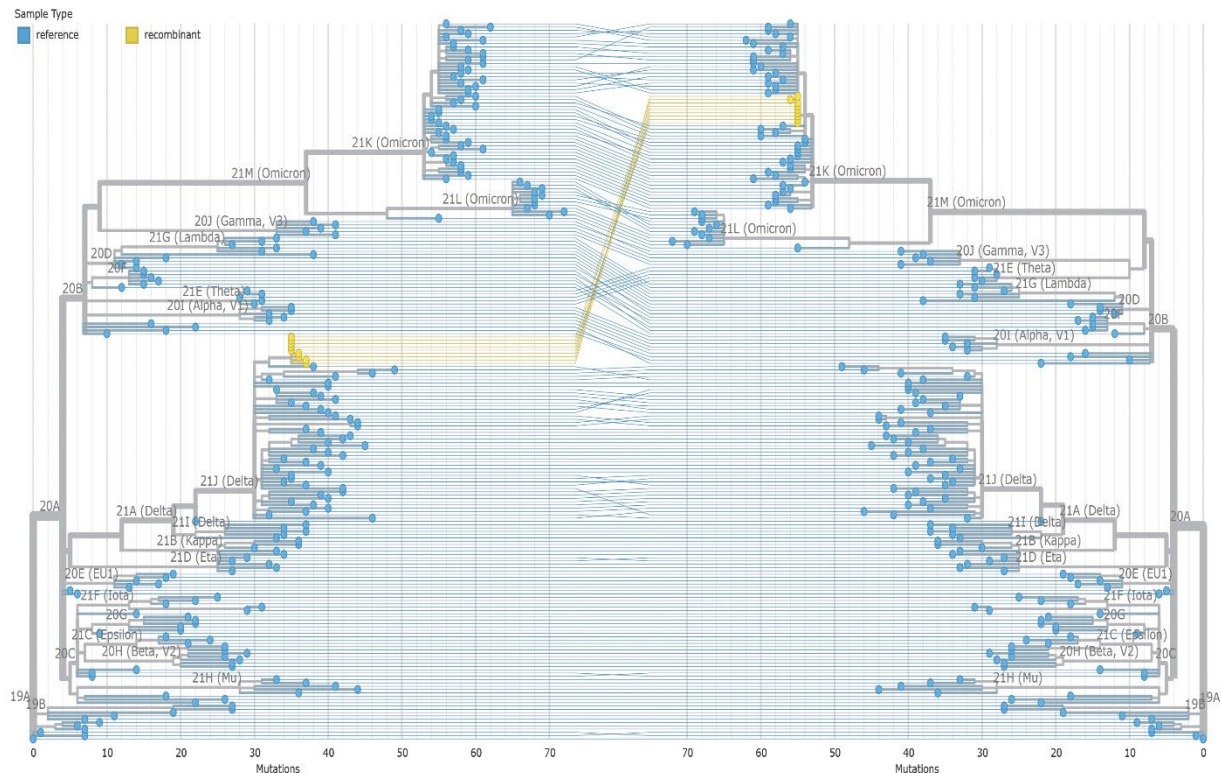


Appendix Figure 1. Composition of candidate recombinant SARS-CoV-2 genomes. A) Amino acid profiles of putative recombinants in the United States and United Kingdom. Delta variant-associated amino acid substitutions are shown in orange and Omicron variant-associated substitutions are shown in purple; hues correspond to the proportion of Omicron or Delta classified sequences that contain that substitution (Appendix, <https://wwwnc.cdc.gov/EID/article/28/7/22-0526-App1.pdf>). Grey boxes indicate

an amino acid substitution common to both Omicron and Delta sequences. White boxes indicate no change relative to Wuhan-Hu-1 virus, and black boxes denote substitutions that are not common to either Delta or Omicron sequences in aggregate. Distinct groups are shown; sequences from the United States appear to have recombination within the spike gene, and samples from 2 clusters from the United Kingdom show recombination upstream of the spike gene (UK cluster 1 represents <https://github.com/cov-lineages/pango-designation/issues/445>, UK cluster 2 represents <https://github.com/cov-lineages/pango-designation/issues/441>). The BA.1.1 (Omicron) deletion associated with spike-gene target failure ($\Delta 69-70$), the receptor-binding domain, and the range containing the recombination location are noted. B) Proportion of reads supporting each single-nucleotide variation and deletion around the recombination site from Illumina (IDT xGen amplicons) datasets generated at the Centers for Disease Control and Prevention. Shown are 2 recombinants (EPI_ISL_8981459, EPI_ISL_8981824), next to a representative AY.119.2 (Delta) genome (EPI_ISL_6811176) and a representative BA.1.1 (Omicron) genome (EPI_ISL_9351600). Each bar shows the proportion of reads containing the given allele (colored by nucleotides A, C, T, and G) at each position for each sample. Asterisks denote deletions. Variants are relative to Wuhan-Hu-1 virus.



Appendix Figure 2. IGV Alignment of Nanopore long reads from one of the recombinant viruses demonstrating the presence of phased Delta 22029-22034 deletion and Omicron T22673C, C22674T, T22679C, C22686T SNVs originating from a single template. Green, red, and blue lines indicate a substitution with respect to Wuhan-Hu-1, while black bars indicate deletions and purple bars are spurious insertions.



Appendix Figure 3. Tanglegram of candidate SARS-CoV-2 recombinants comparing the 5' and 3' sections of the genome. From a reference sequence set (Appendix, <https://wwwnc.cdc.gov/EID/article/28/7/22-0526-App1.pdf>), 2 separate datasets were generated by adding the start (positions 1–22150) and end (22150 on) portion of putative recombinant sample genomes to the references. Two phylogenetic trees (maximum-likelihood) were generated and visualized together in the Nextstrain Auspice tree viewer (12). The 9 putative recombinants are shown in yellow, and the connecting lines between trees pinpoint the corresponding sequences on each tree.